OpenRiskNet

RISK ASSESSMENT E-INFRASTRUCTURE

NanoCommons

Nano-Knowledge Community

Introduction to OpenRiskNet and NanoCommons projects

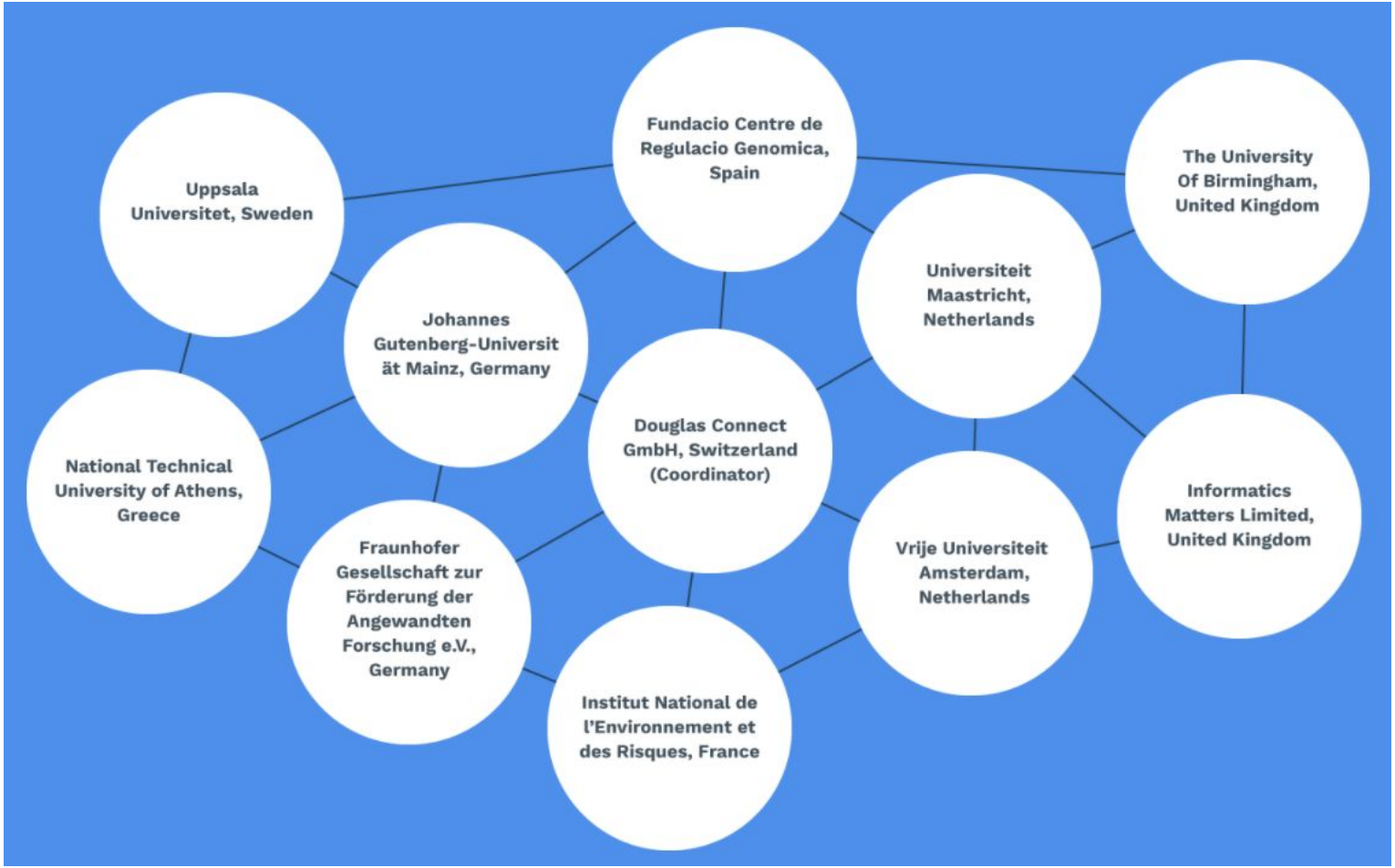
Thomas Exner (Douglas Connect)

U.S. NanoWG meeting 8 March 2018

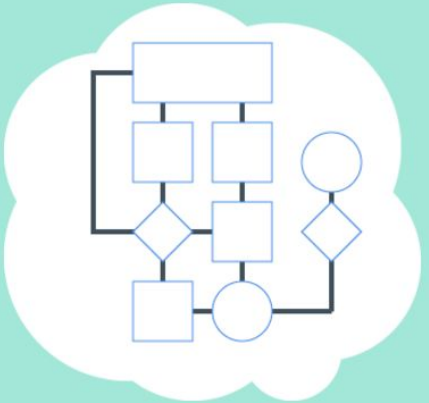
OpenRiskNet: Open e-Infrastructure to Support Data Sharing, Knowledge Integration and in silico Analysis and Modelling in Risk Assessment (Project Number 731075)

OpenRiskNet is a 3 year project (Dec 2016 - Nov 2019) with the main objective to develop an open e-Infrastructure providing **resources and services to a variety of communities requiring risk assessment, including chemicals, cosmetic ingredients, therapeutic agents and nanomaterials**. OpenRiskNet will work with a network of partners, organised within an **Associated Partners Programme**.

Large databases and highly sophisticated methods, algorithms and tools are available for different tasks such as hazard prediction, toxicokinetics, and *in vitro – in vivo* extrapolations to support this transition. However, since these services are developed independently and provided by different groups world-wide, there is **no standardised way to access the data or run modelling workflows**. To overcome the fragmentation of data and tools, OpenRiskNet will provide **open e-Infrastructure resources and services** supporting different scientific communities.


Uppsala Universitet, Sweden
Fundacio Centre de Regulacio Genomica, Spain
The University Of Birmingham, United Kingdom
Johannes Gutenberg-Universität Mainz, Germany
Universiteit Maastricht, Netherlands
National Technical University of Athens, Greece
Douglas Connect GmbH, Switzerland (Coordinator)
Informatics Matters Limited, United Kingdom
Fraunhofer Gesellschaft zur Förderung der Angewandten Forschung e.V., Germany
Vrije Universiteit Amsterdam, Netherlands
Institut National de l'Environnement et des Risques, France

How?



Easily accessible
Standardised
Harmonised
Scalable
Robust
Infrastructure

For whom?

Researchers
Risk assessors
Regulators
Informed public

To what end?

Improve industrial risk assessments
Prototyping new services and apps
Enabled access to integrated resources
Complete and qualified system
Support inovative product development

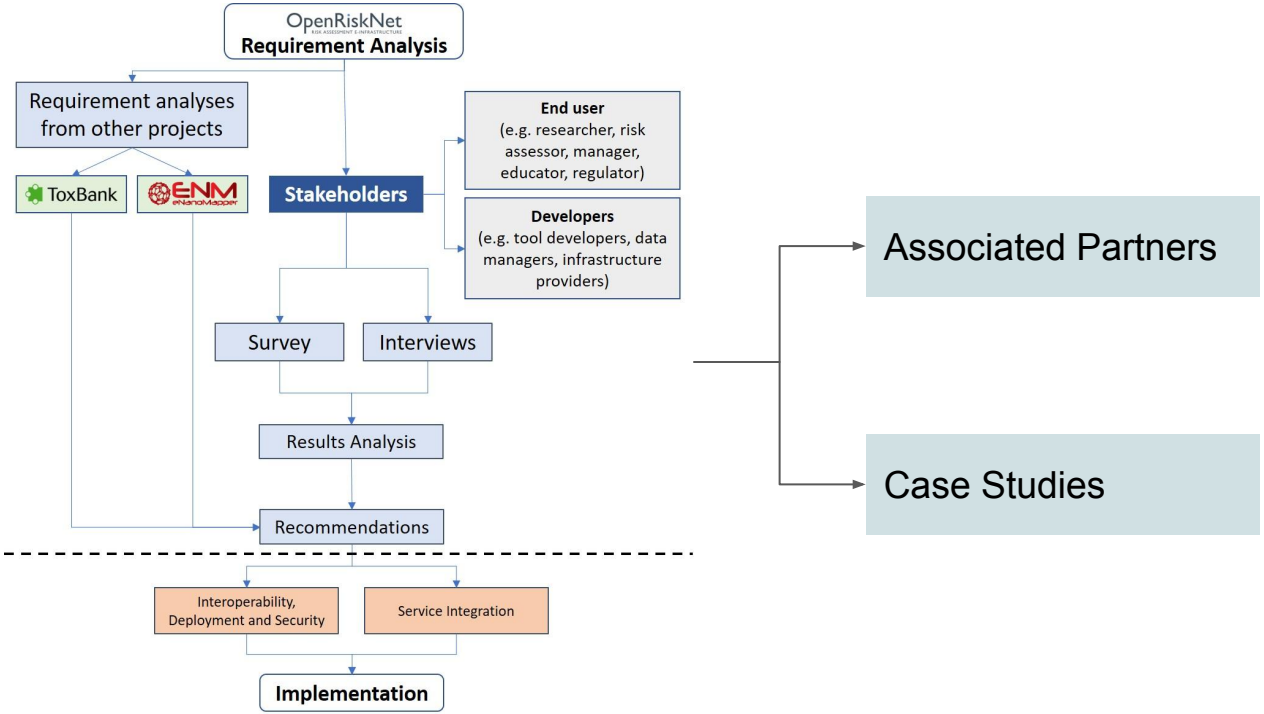
Develop + Deploy Integrated, Secure, and Sustainable e-Infrastructure

Main Components of the OpenRiskNet Infrastructure

1. **Case-study-driven development** - examples of tools to be integrated are selected based on the case study needs
2. Information on case studies in the areas of **chemical and nanomaterial risk assessment** can be found at https://openrisknet.org/development/case-studies/
3. **REST services** providing data and processing/analysis/modelling tools (provided by OpenRiskNet and associated partners)
4. Concept to **harmonise APIs** in an bottom-up approach
5. No strict standards but communication through **semantic interoperability layer**, which provides information on the usage of the data and software including human- and computer readable input/output annotations
6. Microservice architecture based on **containerisation and container orchestration** accompanied by a **discovery service**
7. **Virtual infrastructures**, which can deployed on public or in-house clouds

Requirement Analysis

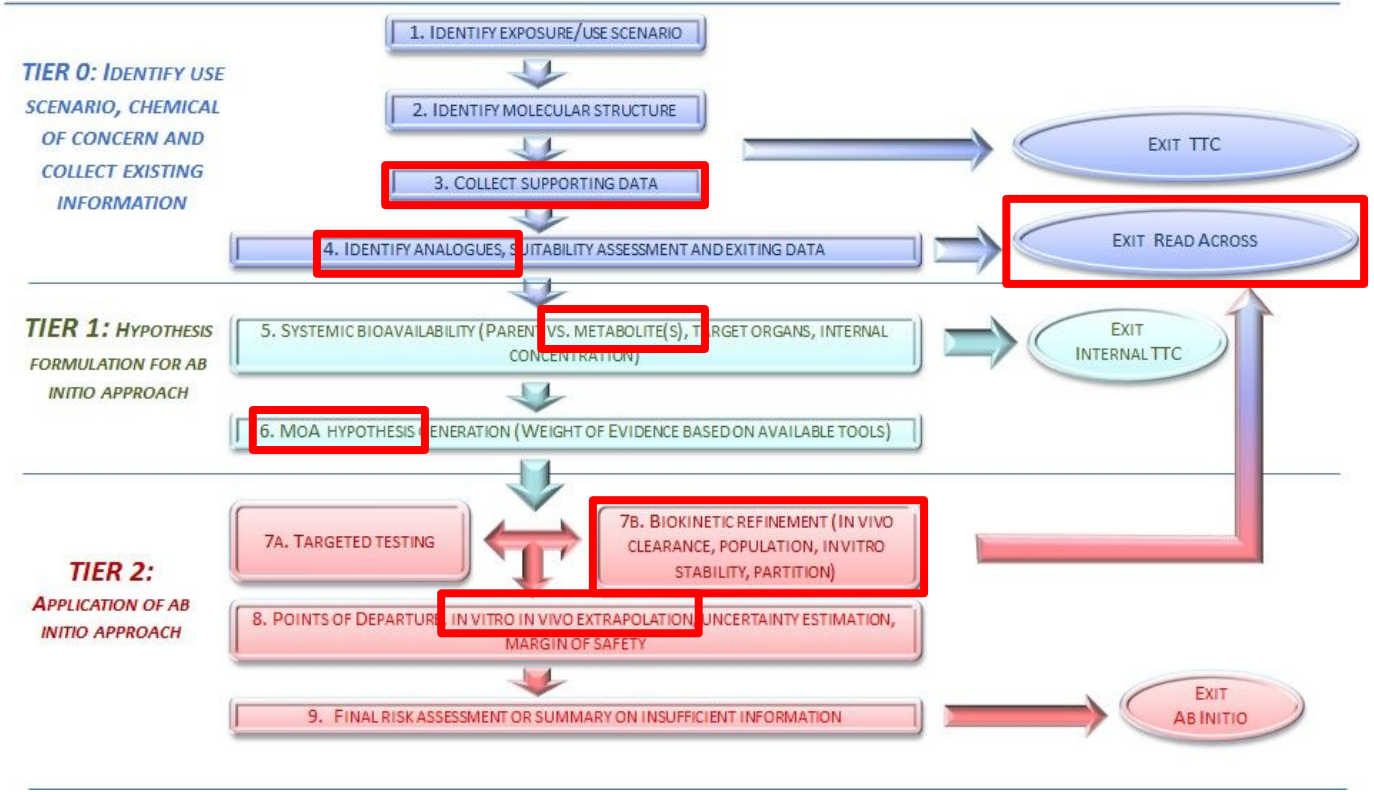
Survey https://goo.gl/forms/kRXfRlpXS0lGvWnp1



Associated Partner Programme https://openrisknet.org/associated-partner-programme/

- The Programme aims at **strengthening the working** ties between the OpenRiskNet Consortium members and other organisations within relevant scientific and technology communities.
- Any organisation such as a university, institute, consortium, non-governmental organisations (NGOs), as well as small and medium enterprises (SMEs) or large commercial companies can become an Associate Partner of OpenRiskNet.
- Types of users:
  - **Service providers** - integrate their databases and software tools into the OpenRiskNet infrastructure
  - **Early adopters** - use the infrastructure for their predictive toxicology and risk assessment tasks
  - **Technology partners** - develop services and tools on which the OpenRiskNet e-infrastructure will be based

Case studies based on risk assessment framework

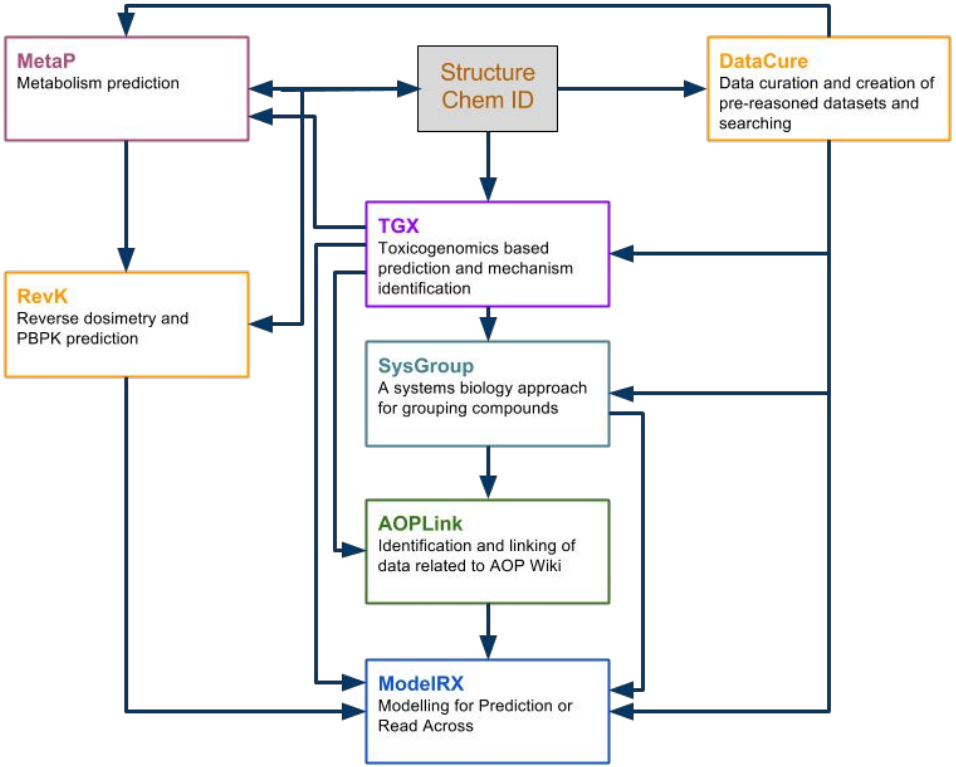


Berggren et al., 2017

Case studies

https://openrisknet.org/development/case-studies/

- **DataCure** - Data curation and creation of pre-reasoned datasets and searching
- **ModelRX** - Modelling for Prediction or Read Across
- **SysGroup** - A systems biology approach for grouping compounds
- **MetaP** - Metabolism Prediction
- **AOPLink** - Identification and Linking of Data related to AOPWiki
- **TGX** - Toxicogenomics-based prediction and mechanism identification
- **RevK** - Reverse dosimetry and PBPK prediction



OpenRiskNet Interoperability, Deployment and Security

- Allowing for interoperability between data and services at the **technical level**
- Providing mechanisms (APIs) for **describing** services and data
- Providing mechanisms for **discovering** relevant services and data
- Defining approaches for **deploying** services as containers
- Allowing for services to be accessed **securely**, as needed by commercial parties
- Allowing for new OpenRiskNet **virtual environments** containing these tools and services to be instantiated by users on cloud or internal architectures
- Providing **federation** of services and security between different OpenRiskNet virtual environments and with external HPC facilities

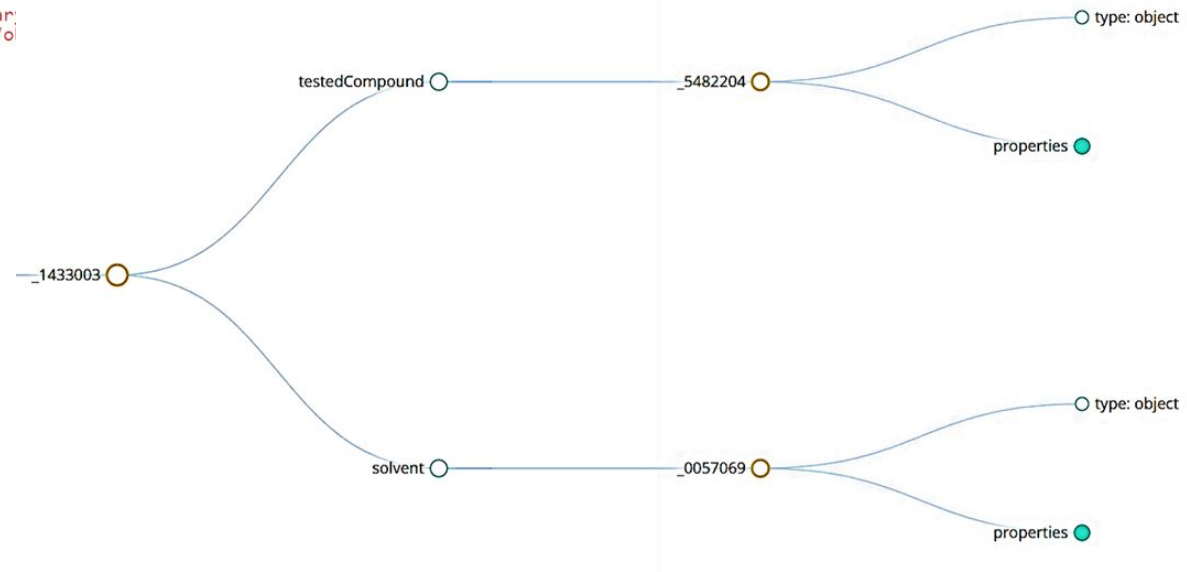
API Design Concept

- Going beyond API concepts realized in previous projects in this area like OpenTox and Open PHACTS is necessary because of two reasons:
  - A **much broader scope of data and tool services** is planned to be integrated and harmonized in OpenRiskNet and
  - **semantic annotation of APIs** needs to be provided giving more information on the service, input and output .
- The approach looking promising to realize the semantic layer is a combination of **OpenAPI definitions** with **json-ld data serialization** to bridge the worlds of **API development and the semantic web**.

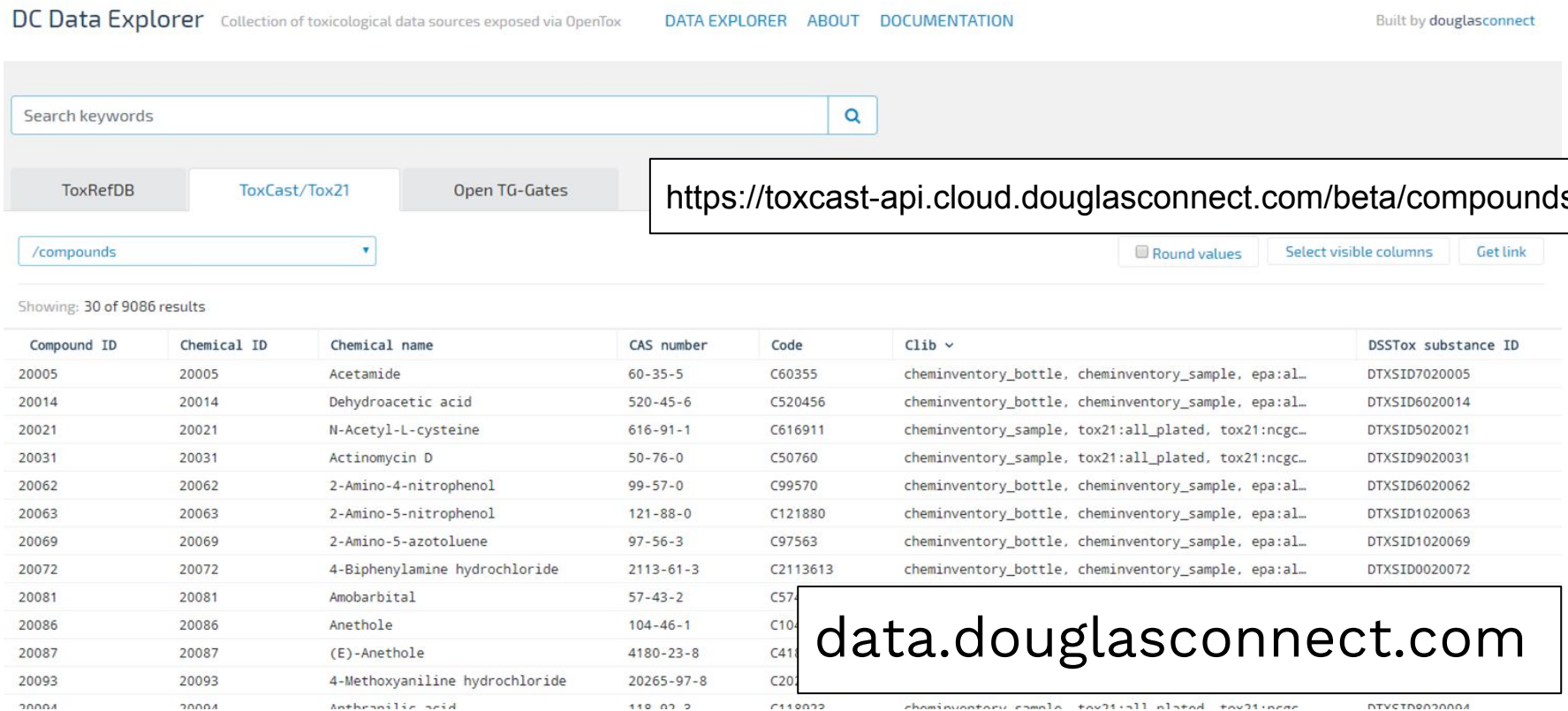
API Design Concept

JSON-LD Input

```
{
  "@context": {
    "smiles": {
      "@id": "http://edamontology.org/format
    },
    "properties": "http://orn.org/properties
    "type": "http://orn.org/type",
    "testedCompound": "http://purl.obolibrar
    "solvent": "http://purl.obolibrary.org/o
  },
  "type": "object",
  "properties": {
    "testedCompound": {
      "type": "object",
      "properties": {
        "smiles": {
          "type": "string"
        },
        "cas": {
          "type": "string"
        }
      }
    },
    "solvent": {
      "type": "object",
      "properties": {
        "smiles": {
          "type": "string"
        },
        "cas": {
          "type": "string"
        }
      }
    }
  }
}
```

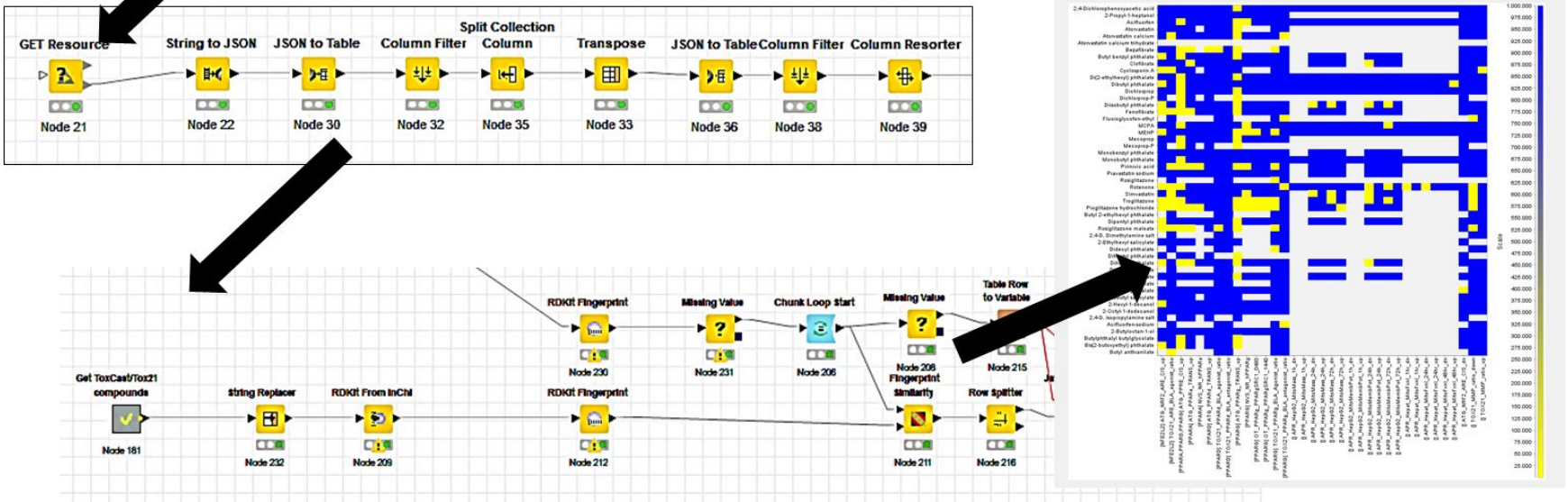


Data API


DC Data Explorer
Collection of toxicological data sources exposed via OpenTox
DATA EXPLORER
ABOUT
DOCUMENTATION
Built by douglasconnect
Search keywords
ToxRefDB
ToxCast/Tox21
Open TG-Gates
https://toxcast-api.cloud.douglasconnect.com/beta/compounds
/compounds
Round values
Select visible columns
Get link
Showing: 30 of 9086 results
Compound ID | Chemical ID | Chemical name | CAS number | Code | Clib | DSSTox substance ID
20005 | 20005 | Acetamide | 60-35-5 | C60355 | cheminventory_bottle, cheminventory_sample, epa:al… | DTXSID7020005
20014 | 20014 | Dehydroacetic acid | 520-45-6 | C520456 | cheminventory_bottle, cheminventory_sample, epa:al… | DTXSID6020014
20021 | 20021 | N-Acetyl-L-cysteine | 616-91-1 | C616911 | cheminventory_sample, tox21:all_plated, tox21:ncgc… | DTXSID5020021
20031 | 20031 | Actinomycin D | 50-76-0 | C50760 | cheminventory_sample, tox21:all_plated, tox21:ncgc… | DTXSID9020031
20062 | 20062 | 2-Amino-4-nitrophenol | 99-57-0 | C99570 | cheminventory_bottle, cheminventory_sample, epa:al… | DTXSID6020062
20063 | 20063 | 2-Amino-5-nitrophenol | 121-88-0 | C121880 | cheminventory_bottle, cheminventory_sample, epa:al… | DTXSID1020063
20069 | 20069 | 2-Amino-5-azotoluene | 97-56-3 | C97563 | cheminventory_bottle, cheminventory_sample, epa:al… | DTXSID1020069
20072 | 20072 | 4-Biphenylamine hydrochloride | 2113-61-3 | C2113613 | cheminventory_bottle, cheminventory_sample, epa:al… | DTXSID0020072
20081 | 20081 | Amobarbital | 57-43-2
20086 | 20086 | Anethole | 104-46-1
20087 | 20087 | (E)-Anethole | 4180-23-8
20093 | 20093 | 4-Methoxyaniline hydrochloride | 20265-97-8
data.douglasconnect.com

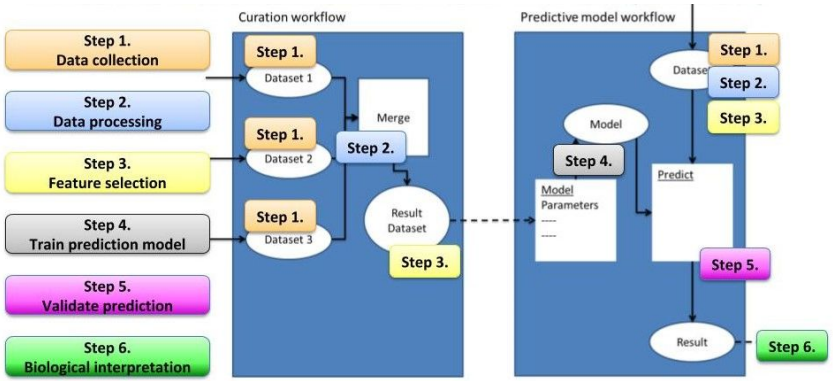
Data API

https://open-tggates-api.cloud.douglasconnect.com/v2/compounds



OpenRiskNet Service Integration

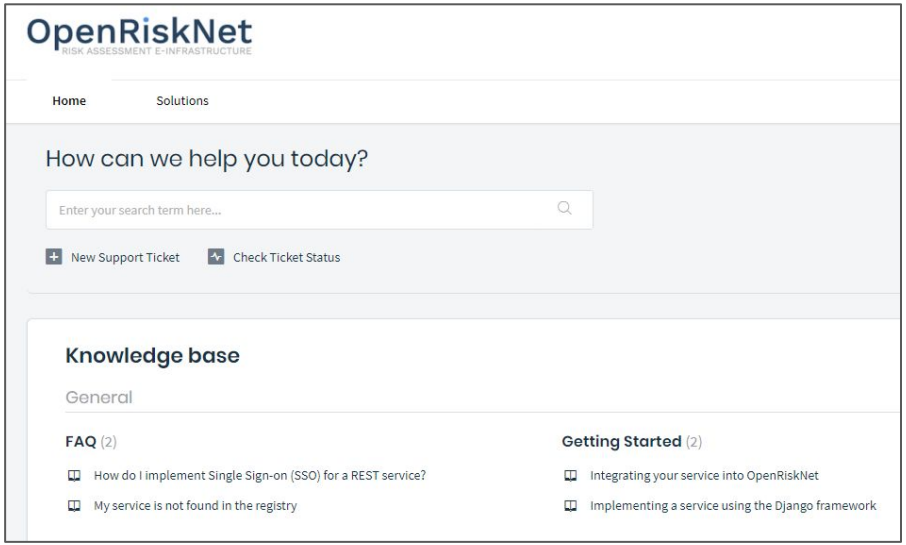
- Integrate a representative number of services as best-practice examples with the corresponding step-by-step documentation
- Harmonisation of approaches to provide information on service status, capabilities and requirements over the interoperability layer
- Adoption of the data schemata and deployment as well as the appropriate authentication & authorisation options
- Guarantee the error-free communication and integration of the services in the infrastructure and provide maintenance of the service



Services requirements for a cell model based prediction

User training and support

- Documentation for the core infrastructure as well as all integrated services
- Training materials, workshops and hackathons
- Maintenance of the services
- User support and help-desk



https://github.com/OpenRiskNet
https://openrisknet.freshdesk.com/

Acknowledgements

OpenRiskNet (Grant Agreement 731075) is a project funded by the European Commission within Horizon2020 Programme

Project partners:

P1 Douglas Connect GmbH, Switzerland (DC)
P2 Johannes Gutenberg-Universitat Mainz, Germany (JGU)
P3 Fundacio Centre De Regulacio Genomica, Spain (CRG)
P4 Universiteit Maastricht, Netherlands (UM)
P5 The University Of Birmingham, United Kingdom (UoB)
P6 National Technical University Of Athens, Greece (NTUA)
P7 Fraunhofer Gesellschaft Zur Foerderung Der Angewandten Forschung E.V., Germany (Fraunhofer)
P8 Uppsala Universitet, Sweden (UU)
P9 Medizinische Universitat Innsbruck, Austria (MUI)
P10 Informatics Matters Limited, United Kingdom (IM)
P11 Institut National De L'environnement Et Des Risques INERIS, France (INERIS)
P12 Vrije Universiteit Amsterdam, Netherlands (VU)

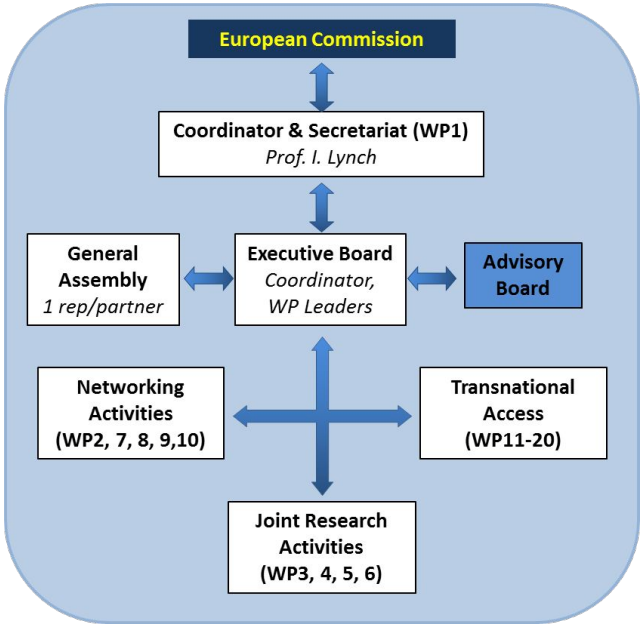
The European Nanotechnology Community Informatics Platform: Bridging data and disciplinary gaps for industry and regulators

NanoCommons
Nano-Knowledge Community

NanoCommons proposes to create an openly accessible e-infrastructure of scientific and cutting edge and managerial excellence provided by a combination of research intensive academic groups and SMEs serving the current and future (unmet) needs of the key research communities and pivotal industrial users and regulators.

NanoCommons will bring pan-European and international added value and innovation opportunities, by answering the increasing demands concerning the prediction of safety of existing and new nanoscale materials for health and environmental sustainability.

4 years project (Jan 2018 - Dec 2021)



NanoCommons
Nano-Knowledge Community

Research infrastructure projects

Joint Research Activities (JRA)

Specific programme to address gaps

Must be innovative research

Research then feeds into NA and TA

All to enhance the field

Networking Activities (NA)

Specific programme to involve the wider communities

Two-way communication

Can include training & community building

To enhance the field

Transnational Access (TA)

Provision of funded access to the infrastructure facilities

Periodic “calls” for access

Managed selection process

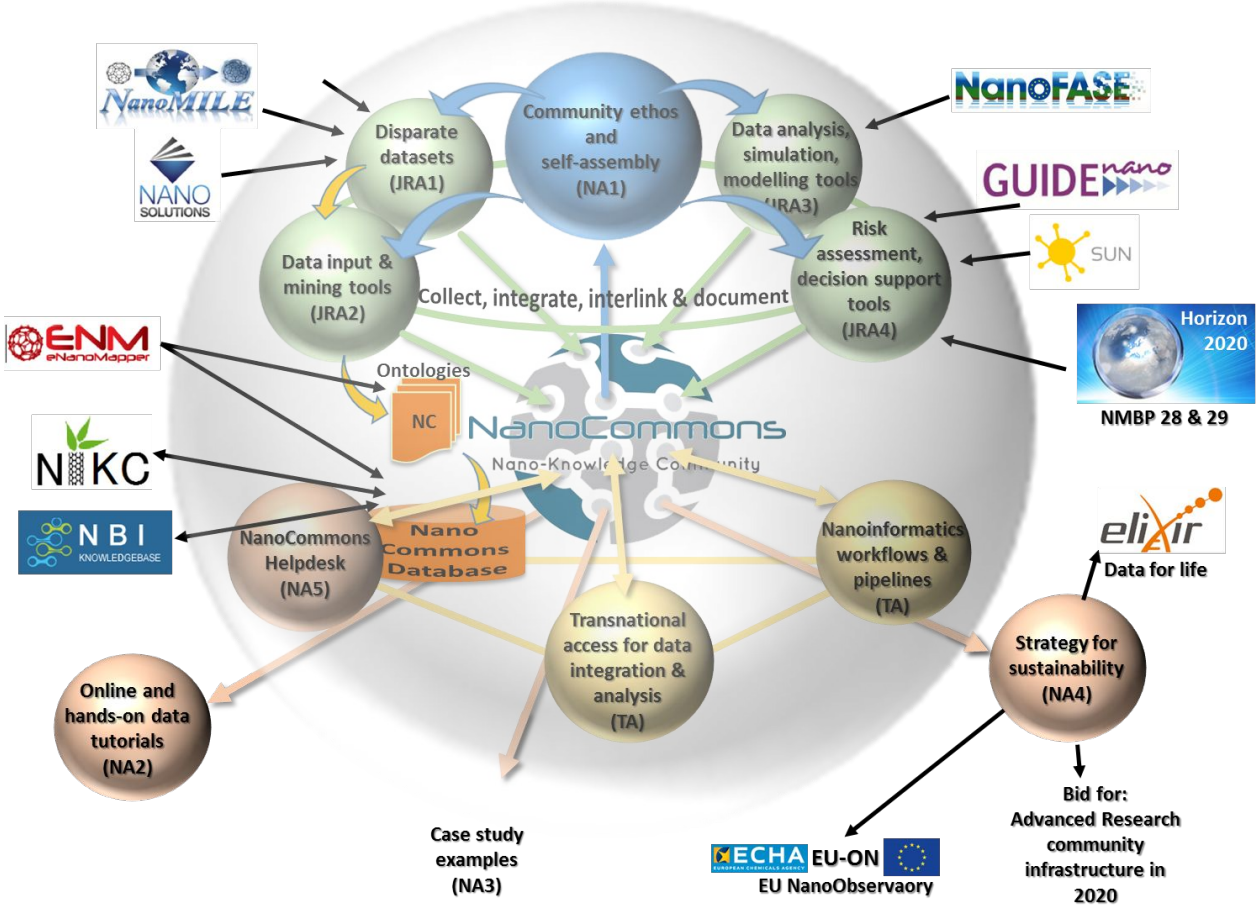
To enhance the field

NanoCommons
Nano-Knowledge Community

NanoCommons objectives

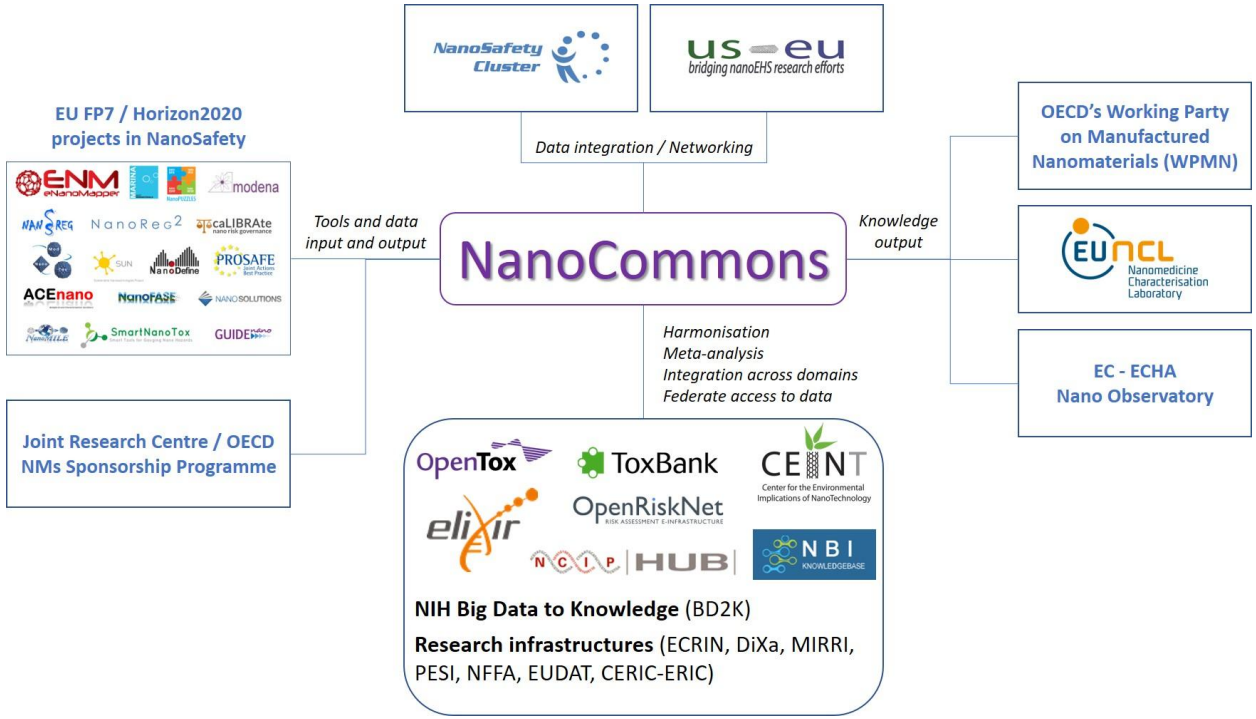
- To develop an **integrated data and methods capture, management and nanoinformatics platform** to enhance the accessibility and reusability of nano-related data and associated protocols (JRA).
- To provide **funded and expert-supported Access** to a range of data / knowledge management tools & nanoinformatics services integrated and developed via the JRA to the widest possible range of stakeholders This will be supported by a helpdesk and through workpackages (WPs) focussed on experimental design and nanoinformatics (TA).
- To continuously assess **community needs** in terms of services required, and align training offers, tools integrated into the TA portfolio and real-life demonstration case studies to address these needs (NA).

NanoCommons
Nano-Knowledge Community



NanoCommons
Nano-Knowledge Community

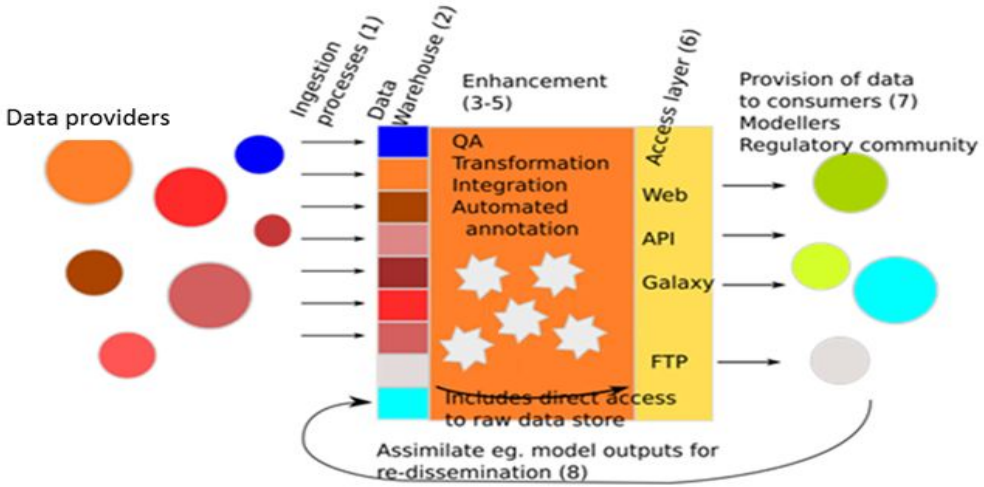
Integration of resources



NanoCommons
Nano-Knowledge Community

NanoCommons Knowledge Hub

1. Ingest data from nanosafety labs
2. Act as a data warehouse, storing all ingested data
3. Apply internal QA, reformatting and homogenisation
4. Develop concepts for data curation
5. Integrate related datasets (for using common or reference NMs)
6. Apply automated annotation pipelines
7. Provide data access via a variety of mechanisms
8. Specifically feed the needs of stakeholders and Users developing new analytical modules
9. Assimilate & disseminate results back to the community
10. Maintain provenance and links at all stages



1- The Ontology exists as a community resource outside this model
2- The knowledge enhancement layer is new. All partners will participate in enhancing this

NanoCommons
Nano-Knowledge Community

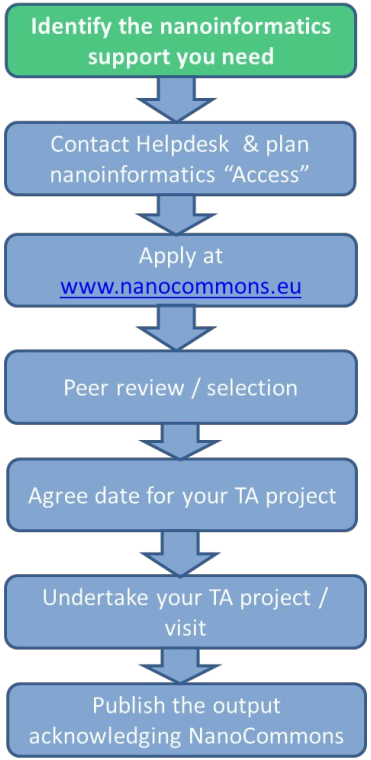
Knowledge infrastructure development

- **Ontology development and integration** - NanoCommons will build on efforts from eNanoMapper, NanoMILE, CEINT NIKC, SmartNanoTox, NanoFASE, and ACEnano to develop nano-specific ontologies to ensure nano-relevant terms are in place for supporting templates and data processing tools.
- **Data warehousing** - store all raw data and processed data generated according to protocols defined in JRA activities by users; provide support and processes for preparing and inputting datasets, templates and protocols into the system.
- **Tools for data search and retrieval** - develop tools to allow a variety of searches across the NanoCommons KnowlegdeBase including all integrated data sources by substance identifier, experimental measurements, and/or chemical or biological identity.
- **Text mining and data curation** - develop tools for automatic text searching and deriving high-quality information.
- **Service descriptions and discover service** - provide a registry for all tools and data service integrated in the NanoCommons infrastructure; services will be described in detail with the expected input and output formats, references to the scientific as well as technical documentation and sorted per the types of data to which the services are applicable.

NanoCommons Transnational Access (TA)

Access for users to:

- KnowledgeBase integrating data & tools
- Ontology services
- Data processing & import services
- Data mining tools
- Omics metadata support
- Predictive tools
- Visualisation / Risk tools
- Nanoinformatics Workflows for key experiments



NanoCommons
Nano-Knowledge Community

Acknowledgements

NanoCommons (Grant Agreement 731032) receives funding from the European Union Horizon 2020 Programme

P1 - UoB - The University of Birmingham
P2 - DC - Douglas Connect
P3 - NERC CEH - Natural Environment Research Council
P4 - NTUA - National Technical University of Athens
P5 - NUID UCD - National University of Ireland Dublin - University College Dublin
P6 - LEITAT - ACONDICIONAMIENTO TARRASENSE ASSOCIACION
P7 - BfR - Bundesinstitut für Risikobewertung
P8 - BIONANONET - BioNanoNet
P9 - PLUS - Universitat Salzburg / Paris Lodron University of Salzburg
P10 - NM - NovaMechanics Ltd
P11 - BIOMAX - BIOMAX Informatics AG
P12 - UM - Universiteit Maastricht
P13 - DU - Duke University
P14 - OSU - Oregon State University

NanoCommons
Nano-Knowledge Community